

Análisis práctico de folksonomías

El caso de los gestores bibliográficos sociales



Nadina Yedid

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Departamento de Bibliotecología / nadineyedid@hotmail.com

Resumen

Se presenta un análisis de las folksonomías desarrolladas en los gestores bibliográficos sociales, como medio para identificar el potencial uso de las mismas en la recuperación de información en entornos digitales poco controlados. Se propone también identificar métodos y técnicas no contempladas actualmente en esos sitios web, para mejorar la calidad de las folksonomías en cuanto herramientas para la recuperación de la información. A tales fines, se estudian las folksonomías desarrolladas en los gestores bibliográficos CiteULike, Mendeley y Bibsonomy. Se detallan sus principales características en cuanto a tipo, forma y construcción de las etiquetas. Se describen también las interfaces de creación y edición de etiquetas en cada uno de estos gestores. Se concluye que las folksonomías pueden resultar útiles para la recuperación de información en los gestores bibliográficos, en función de la forma y estructura con la que son creadas, y se proponen nuevos métodos y técnicas para mejorar la calidad de las mismas.

Palabras clave

Folksonomías
Indización
Etiquetas
Gestores bibliográficos
CiteULike
Mendeley
Bibsonomy

Abstract

Practical analysis of folksonomies: the case of the reference management software.

An analysis of the folksonomies developed in reference management software is performed, as a way to identify their potential use for the information retrieval in low controlled digital environments. There is also a proposal to identify methods and techniques not currently used in those websites, to improve the quality of the folksonomies as information retrieval tools. For that reason, the folksonomies developed at the reference management software CiteULike, Mendeley and Bibsonomy are studied. Their principal characteristics regarding type, form and construction of the tags are detailed. The interfaces used for creation and edition of the tags in each one of the reference management software are also described. The conclusion is that, based on the form and structure used to create them, the folksonomies might be useful

Keywords

Folksonomy
Indexing
Tags
Reference management software
CiteULike
Mendeley
Bibsonomy

for information retrieval at this kind of website, and new methods and techniques are proposed to improve the quality of the folksonomy tags.

Artículo recibido: 19-05-2016. Aceptado: 01-09-2016

Introducción

Al intentar recuperar los orígenes de la indización y la clasificación nos podemos remontar a los confines de la historia. Parece ser que, como proponen muchos teóricos, los seres humanos estamos biológicamente impelidos a organizar de una u otra forma el mundo que nos rodea. Desde las categorías aristotélicas hasta las actuales ontologías de la web semántica, ha habido un continuo empeño por ordenar nuestro entorno, un paradigma tras otro. Según David Weinberger (2007) el mundo comenzó siendo misceláneo, pero no se quedó así. Las personas organizan absolutamente todo el mundo a su alrededor, siguiendo distintos estándares de clasificación. La forma en que organizan los papeles personales y los archivos de la computadora, pero también la ropa en el armario, la vajilla en la alacena y los víveres en la heladera, todo responde a una forma de comprender el mundo en que se desenvuelven, y que les permite manejarse con eficiencia y seguridad.

De acuerdo con Arlene Taylor y D. N. Joudrey (2009) la capacidad de aprendizaje de los humanos se encuentra íntimamente relacionada con su habilidad para analizar, organizar y recuperar la información y el conocimiento; reconociendo patrones, comparando experiencias, conceptos e ideas y buscando relaciones entre todos ellos. Para esta autora, las personas organizan porque de esa forma pueden tener un mejor conocimiento del mundo que las rodea. Organizan porque eso les permite ahorrar tiempo y, sobre todo, porque eso les facilita recuperar la información cuando la necesiten nuevamente.

Refiriéndonos específicamente al ámbito de la Bibliotecología y la Documentación, el desarrollo de lenguajes documentales para la organización de la información registrada comenzó hacia finales del siglo XIX. Imbuido de la mentalidad positivista, a partir del año 1876 comenzaron los primeros aportes de Cutter relacionados con los objetivos de los catálogos. A estos le siguieron los trabajos de Dewey en su clasificación decimal y la posterior clasificación decimal de Otlet y Lafontaine; la clasificación bibliográfica de Bliss; la clasificación facetada de Ranganathan; y los aportes de Taube y Mooers a las listas de vocabularios controlados. Los distintos trabajos mencionados permitieron conformar un corpus de conocimientos teóricos que funcionaron como la base para el desarrollo de diferentes métodos de organización de la información (Martínez Tamayo y Valdez, 2009).

Sin embargo, más allá de todos los métodos y sistemas desarrollados, siempre existió entre los profesionales de la información una contraposición entre aquellos que abogan por la utilización de alguno de estos métodos para la organización documental y quienes propugnan el uso del lenguaje natural como principal fuente para la recuperación de información.

Mario Barité (2014) retoma y extiende la separación en fases propuesta por Chu en relación a la dicotomía entre la organización de la información basada en el lenguaje natural y la organización a partir de un lenguaje documental. De acuerdo a esta

diferenciación, se pueden percibir cuatro distintas fases, siendo la última de ellas la correspondiente al Siglo XXI, en el que predominan los modelos de caracterización ontológica para la web semántica, y el uso del lenguaje natural, asociado al desarrollo de modelos cooperativos accionados por los usuarios en entornos web 2.0.

Es posible interpretar que dentro de este último grupo se puede ubicar el fenómeno que ha tenido lugar en los últimos años, conocido con el nombre de “Folksonomy” o “Folksonomías”. Sin embargo, como se intentará demostrar a lo largo de este trabajo, las folksonomías no son solo la aplicación del lenguaje natural en la organización de la información. Tal como propone Barité (2014), los sitios web que funcionan como plataformas para el desarrollo de folksonomías han evolucionado en sus técnicas de control y corrección de etiquetas, y es posible pensar en métodos que estructuren la construcción de las etiquetas al punto tal que podría llegar a considerarse a las folksonomías como vocabularios semicontrolados.

El presente trabajo se propone entonces conocer cuáles son las principales características de las etiquetas que componen una folksonomía y qué métodos se pueden implementar para mejorar su funcionalidad como herramientas para la recuperación de información. La hipótesis central que guía el trabajo se basa en la idea de que las folksonomías pueden constituir lenguajes propicios para la recuperación de información en entornos poco controlados, con altos niveles de crecimiento, y funcionales respecto a la participación de los usuarios.

Marco teórico

Tipos de folksonomías

Al repasar la bibliografía existente sobre folksonomías es posible apreciar que gran cantidad de teóricos han propuesto esquemas de diferenciación o tipologías de folksonomías, basados en diferentes aspectos de las mismas. A los fines prácticos de este trabajo, se repasan a continuación dos de las tipologías más citadas en los estudios sobre la temática de interés, a saber, la tipología de Golder y Huberman (2006), y la tipología de Cañada.

En el esquema de Golder y Huberman (2006) se distinguen siete tipos de etiquetas, según la función que cumplen:

1. Etiquetas que identifican sobre qué (o quién) trata el recurso, es decir, cuál es el tema o su contenido.
2. Etiquetas que identifican qué es el recurso, es decir, que tipo de cosa es el contenido (por ejemplo un blog, o un artículo).
3. Etiquetas que identifican quién es el creador del recurso.
4. Etiquetas que funcionan como categorías refinadoras, es decir, que no pueden utilizarse solas, sino que acompañan a otras etiquetas para especificarlas o refinarlas.
5. Etiquetas que identifican cualidades o características del recurso, adjetivos tales como gracioso, tonto, inspirativo, o que dan cuenta de la opinión del usuario sobre el contenido.
6. Etiquetas de auto referencia, es decir, etiquetas que inician con “mi”, tales como “miscomentarios”, o “micontenido”, que demarcan la relación de los contenidos con el usuario.
7. Etiquetas utilizadas para la organización de tareas, como por ejemplo “paraleer”, o “parabuscarttrabajo”.

De acuerdo con estos autores, las primeras cuatro funciones pueden servir para generar etiquetas útiles para toda la comunidad de usuarios mientras que las últimas tres no proporcionan ningún beneficio para el resto de la comunidad.

Con una orientación similar, basada en la utilidad que pueda tener una etiqueta para la comunidad de usuarios de un sitio web, Cañada (2016) propone otra clasificación de las folksonomías. Para este autor la motivación que puedan tener los usuarios al momento de seleccionar las etiquetas que utilizarán para describir los contenidos puede influir en la eficiencia del sistema para el colectivo de usuarios. Existen distintos estilos de etiquetado y los entornos pueden funcionar mejor o peor de acuerdo al estilo de etiquetado que predomine en ellos:

1. Etiquetado egoísta: se utilizan como etiquetas términos que solo sirven para poder volver a encontrar aquello que se está etiquetando. Se utilizan etiquetas personales, muy significativas para la persona, pero carentes de significado para el resto de los usuarios. Con el tiempo algunos usuarios tienden a sistematizar este estilo utilizando etiquetas más temáticas que pueden aumentar el beneficio social (como por ejemplo “divertido”) y menos etiquetas de acciones o de términos muy personales que llenan de ruido el sistema.
2. Etiquetado amiguista: se etiqueta aquello que se desea compartir con personas del propio círculo (amigos, familia, etc.) y se utilizan etiquetas con significado para el grupo, pero desconocido para el resto de los usuarios, por ej “cadius” “asobine”. Este tipo de etiquetado resulta poco beneficioso para grupos grandes, pero puede resultar altamente beneficioso en grupos pequeños.
3. Etiquetado altruista: se etiqueta un contenido con la intención de que este pueda ser compartido con otros usuarios, y por lo tanto se utilizan etiquetas que se consideran con mayor poder de descripción, conocidas o aceptadas. Este tipo de etiquetado presenta un beneficio social muy alto ya que facilita la recuperación de información por parte del resto de los usuarios del sistema.
4. Etiquetado populista: se etiqueta un contenido con términos populares y llamativos con miras a aumentar su atractivo y hacer que obtenga más visitas. Este etiquetado no conlleva ningún tipo de beneficio social.

Problemas en el empleo de las folksonomías y jardinería de etiquetas

Tal como se ha explicado en el apartado anterior, una de las principales características de las folksonomías es que no utilizan ningún vocabulario controlado para su conformación, sino que se encuentran basadas en el lenguaje natural y las palabras claves. Como consecuencia de ello, las folksonomías padecen los mismos inconvenientes que el resto de sistemas de indización asentados sobre el lenguaje natural que muestran muchas dificultades para superar la ambigüedad, la sinonimia, la polisemia, la homonimia y el problema de la variación del llamado “nivel básico”.

Sin embargo, de acuerdo con diferentes estudios realizados por distintos autores como Guy y Tonkin (2006) y Spiteri (2007), la variabilidad en las etiquetas no se relaciona simplemente con los problemas de polisemia y sinonimia, sino también con el hecho más básico de que no existen reglas respecto de cómo construir las etiquetas. Es decir, no existen guías respecto de cómo manejar las formas plurales y singulares, la puntuación, el orden de las palabras, entre otras cuestiones. Existen etiquetas que representan palabras conjugadas, o “palabras” compuestas de muchas palabras, o incluso palabras inventadas que solo tienen significado para un grupo particular de personas.

Para muchos autores, estos problemas podrían mejorar a partir de la implementación de una serie de buenas prácticas en las plataformas que soportan las folksonomías. La definición de estas buenas prácticas ha tomado el nombre de “tag gardening” o jardinería de etiquetas, y es uno de los problemas más estudiados en relación con las folksonomías. Guy y Tonkin (2006) proponen una lista de buenas prácticas que incluye:

1. Utilizar siempre formas plurales en lugar de singulares.
2. Utilizar letras minúsculas.
3. Agrupar las palabras utilizando un guión bajo.
4. Seguir las convenciones iniciadas por otros, y agregar sinónimos.

Así mismo, los autores reconocen que sería posible realizar mejoras desde el propio software a partir de la utilización de métodos de chequeo de errores tipográficos u ortográficos, y de la sugerencia de etiquetas ya existentes en el sistema. El estudio de sistemas de recomendación de etiquetas es uno de los más avanzados en la actualidad, existiendo pruebas y estudios de caso, que proponen y evalúan diferentes algoritmos de recomendación. De acuerdo con Jäschke et al. (2007) los sistemas de recomendación pueden servir a diferentes propósitos, tales como aumentar las posibilidades de que un recurso sea etiquetado, recordar a un usuario de qué trata el recurso, y consolidar el vocabulario entre los usuarios. Por su parte, Font et al. (2015) proponen que un sistema de recomendación de etiquetas debiera traer como beneficios aparejados la consolidación de un vocabulario convergente entre los usuarios, el aumento de la calidad de las etiquetas, y la disminución del costo en el proceso de etiquetado por parte de los usuarios.

Hammond et al. (2005) analizan las posibilidades de desarrollar un software que envíe carteles de ayuda como los siguientes: “sugiere etiquetas para mí”, “encuentra sinónimos automáticamente”, “ayúdame a usar etiquetas ya utilizadas por otros”, “infiere jerarquía a partir de la etiqueta”, y “facilita el ajuste de las etiquetas a los contenidos antiguos”.

Si bien la propuesta de utilización de buenas prácticas parece estar calando hondo en la actualidad, muchos autores miran con recelo esta forma de normalización de las folksonomías. En este sentido, Montero (2006) propone que no se debe olvidar que una de las principales razones del éxito del etiquetado es precisamente el bajo esfuerzo cognitivo que implica la asignación de etiquetas libres, sin tener que amoldarse a reglas o esquemas predefinidos. Mathes (2004) coincide con esta línea de pensamiento al señalar que el motivo por el que las folksonomías se convirtieron en un fenómeno es porque permiten que usuarios sin ningún tipo de entrenamiento, y con muy poco esfuerzo cognitivo, puedan participar en el sistema.

Aspectos metodológicos

El alcance del estudio se ha limitado al examen de los gestores bibliográficos sociales, como exponentes de ámbitos web no controlados y de crecimiento acelerado a través de la participación de los usuarios. La muestra de etiquetas analizada se acotó al conjunto integrado por tres gestores bibliográficos sociales: CiteULike, Mendeley y Bibsonomy. Dichas etiquetas fueron consideradas desde el grado de utilidad que pueden presentar para la comunidad, así como también desde los aspectos lingüísticos que intervienen en su construcción y la forma en la que son construidas. No se realizaron comparaciones a nivel semántico (establecimiento de sinónimos y cuasi-sinónimos).

La composición de la muestra se obtuvo a partir de realizar un proceso de búsqueda y recuperación de información en cada uno de los gestores bibliográficos seleccionados: se realizaron un total de cinco búsquedas de información en cada gestor, cuyos términos de búsqueda fueron (respectivamente) “Mass media”, “Graphene”, “Biomedicine”, “Java”, y “Anthropology”. Con dichas búsquedas se pretendió incluir diferentes áreas disciplinares como son las Ciencias Sociales, las Ciencias Exactas, las Ciencias Naturales, las Ciencias de la Computación y las Humanidades. Para cada una de las búsquedas de información realizadas, se consultaron diez de los registros obtenidos como resultado y se recuperaron las palabras claves que contenía cada uno de ellos. Las etiquetas fueron almacenadas en un archivo Excel poniendo especial cuidado en identificar a qué registro pertenecía cada etiqueta, así como también en mantener el uso de mayúsculas y minúsculas, signos diacríticos, símbolos y caracteres de alfabetos no latinos. Como resultado del proceso de muestreo se obtuvo una base de datos compuesta por las etiquetas de ciento cincuenta registros totales (cincuenta por cada gestor bibliográfico), que suman un total general de 1.306 etiquetas.

La base de datos fue sometida a estudio a partir de una matriz de análisis diseñada ad-hoc para la presente investigación. La misma se compone de 3 secciones, que incluyen distintas cantidades de variables a observar. Las primeras 2 secciones se orientan a identificar el primer objetivo específico de esta investigación, es decir, cuáles son las características de las folksonomías estudiadas. La última sección se dirige a indagar sobre el segundo objetivo de la investigación, relacionado con conocer los métodos de generación y corrección de etiquetas utilizados en estos gestores, con miras a proponer mejoras en ese campo de acción.

Las secciones e indicadores que integran la matriz de análisis son:

Sección 1. Clasificación de las etiquetas. Se incluyen en esta sección los indicadores empleados por dos investigaciones analizadas en los fundamentos de este trabajo. Por un lado, la clasificación propuesta por Golder y Huberman (2006), que divide a las etiquetas en siete tipos posibles. Por otro lado, se incluyen también indicadores provenientes de la clasificación propuesta por Cañada (2016), quien divide a las etiquetas en cuatro grupos generales. Para el análisis de etiquetas en base a la clasificación de Golder y Huberman se incluyen los siete indicadores originalmente propuestos por los autores, más un octavo indicador adicionado ad-hoc para la presente investigación. Este consiste en una categoría denominada “Otra”, que agrupará aquellas etiquetas que por motivos variables no puedan ser incluidas dentro de alguna de las siete categorías originales. Para el análisis de etiquetas según la clasificación de Cañada, se propone una variación que reúne dos categorías originales (la categoría “Etiquetas egoístas” y “Etiquetas amiguistas”) en una única categoría a llamarse “Etiquetas egoístas/amiguistas”. Esto en función de que resulta imposible saber si una etiqueta tiene sentido para un único usuario en particular, o bien si también tiene sentido para su grupo de pertenencia. Además, se considera que a los fines de esta investigación resultará suficiente con conocer si las etiquetas contribuyen a la recuperación de información por parte de toda la comunidad de usuarios (“Etiquetas altruistas”) o no lo hacen (“Etiquetas egoístas/amiguistas”). Se mantiene además la cuarta categoría del esquema original, denominada “Etiquetas populistas”.

Sección 2. Construcción, forma y escritura de las etiquetas. La primera parte de esta sección apunta a conocer cuáles son las construcciones lingüísticas que componen las etiquetas. A tales fines, se consignan doce indicadores que permitirán determinar si las etiquetas corresponden a adjetivos, adverbios, conjunciones, deícticos, frases no sustantivadas, frases sustantivadas, nombres, preposiciones, sustantivos, verbos en infinitivo, verbos conjugados, u otra forma lingüística. Se espera que dichas categorías permitan analizar si las etiquetas se componen de una estructura funcional para la

expresión de conceptos, o bien si utilizan términos a los que no puede atribuírsele un significado conceptual concreto y cuyo uso resulta ambiguo e inespecífico.

La segunda parte de esta sección se compone de seis indicadores dirigidos a medir el nivel de variación en las etiquetas, para aquellas etiquetas que representan exactamente los mismos conceptos. Estos parámetros apuntan a conocer si se generan diferentes etiquetas en función de cuestiones tales como: uso de la forma plural y de la forma singular; uso de mayúsculas y de minúsculas; uso (o no) de espacios y guiones; y traducciones de palabras a otros idiomas.

Sección 3. Análisis de métodos de generación y corrección de etiquetas por gestor. La última sección de la matriz de análisis apunta a conocer cuáles son los métodos que se utilizan en cada uno de los gestores bibliográficos analizados para la creación y corrección de etiquetas por parte de los usuarios. Se proponen para ello cinco indicadores básicos que intentan identificar las buenas prácticas más recomendadas en la bibliografía, con miras a discernir si estas prácticas son utilizadas en los gestores estudiados. Esas prácticas incluyen la posibilidad de visualizar la nube de etiquetas y otras etiquetas asignadas al mismo recurso, la sugerencia de etiquetas por parte del sistema, el ofrecimiento de instrucciones claras y simples para la generación de etiquetas, y la autocorrección de etiquetas por parte del sistema. Se incluye además un último indicador de carácter abierto que permita realizar observaciones generales respecto a este tema.

Resultados obtenidos

Clasificación de las etiquetas

Clasificación a partir del esquema de Golder y Huberman.

Al aplicar la primera sección de la matriz de análisis a la muestra de etiquetas, se obtiene una aproximación al tipo de etiquetas utilizadas en cada uno de los gestores bibliográficos analizados. La siguiente tabla (Tabla 1) permite observar la cantidad total de etiquetas que componen la muestra de cada gestor, y la cantidad correspondiente a cada una de las siete categorías propuestas en la clasificación de Golder y Huberman. La tabla muestra también una octava categoría, adicionada ad-hoc para esta investigación, que incluye etiquetas que por diferentes motivos no han podido ser ubicadas en ninguna de las siete categorías propuestas por el esquema original de clasificación.

Tal como se puede apreciar en la Tabla 1, resulta evidente que en los tres gestores analizados, las etiquetas de contenido son las más utilizadas por los usuarios. Los siguientes gráficos permiten comprobar que, más allá de la variabilidad en la cantidad absoluta de etiquetas (relacionada con el total de etiquetas de cada gestor), se observa una constante en el porcentaje de las etiquetas de contenido, que ronda el entre el 57% y el 67% del total de etiquetas (Gráficos 1, 2 y 3).

Un grado de similitud semejante se puede encontrar también al observar los porcentajes de los seis tipos de etiquetas restantes. Para el gestor CiteULike, las etiquetas más utilizadas luego de las de contenido son las etiquetas refinadoras, y luego las organizativas, con un 9% y un 7% de aparición, respectivamente. Se observa también un porcentaje importante, del 12%, de etiquetas que fueron agrupadas dentro de la categoría “otra”. Por su parte, en el gestor Mendeley también se observa que las etiquetas refinadoras y organizativas son las siguientes en importancia, pero en un orden invertido. Es decir, las etiquetas organizativas representan el 10% de la

CiteULike		Mendeley		Bibsonomy	
Tipo de etiqueta	Cantidad	Tipo de etiqueta	Cantidad	Tipo de etiqueta	Cantidad
Autoreferencial	2	Autoreferencial	7	Autoreferencial	4
Contenido	357	Contenido	271	Contenido	200
Creador	10	Creador	11	Creador	0
Cualitativa	6	Cualitativa	7	Cualitativa	4
Organizativa	39	Organizativa	47	Organizativa	15
Otra	61	Otra	92	Otra	32
Refinadora	46	Refinadora	24	Refinadora	43
Tipo de recurso	8	Tipo de recurso	15	Tipo de recurso	5
Total de etiquetas	529	Total de etiquetas	474	Total de etiquetas	303

Tabla 1: Tipos de etiquetas por gestor. Fuente: elaboración propia

muestra, mientras que las refinadoras representan el 5% de la muestra. También aquí se observa la alta presencia, del 19%, de etiquetas que no pudieron ser clasificadas de acuerdo al esquema original. Finalmente, el gestor Bibsonomy retoma la distribución observada en CiteULike, viéndose más representadas las etiquetas refinadoras y luego las organizativas, con un 14% y un 5% de aparición, respectivamente. Las etiquetas categorizadas como “otras” suman el 11% de la muestra.

Clasificación de etiquetas en CiteULike

Al analizar específicamente las etiquetas de este gestor, es posible apreciar que todos los tipos de etiquetas propuestas por la clasificación de Golder y Huberman se ven representados en la muestra. Tal como se expresó en el apartado anterior, las etiquetas de contenido representan la mayor parte de la muestra, alcanzando un 67% del total de etiquetas estudiadas. Estas etiquetas se presentan en formas variables: muchas veces se componen de palabras claramente relacionadas con la temática de los documentos, tales como “Java” o “Software”. Otras veces, se trata de palabras abreviadas o modismos propios del usuario o de la comunidad académica, pero que aun así pueden ligarse directamente con el contenido del documento, como por ejemplo “JVM” (por “Java Virtual Machine”).

Resulta interesante destacar que todos los registros analizados para este gestor presentan al menos una etiqueta de contenido, dentro del conjunto de etiquetas. Esto implica que todos los registros cuentan con por lo menos una voz que facilita su recuperación por la temática tratada en el mismo. Así mismo, es posible detectar que la mayoría de los registros contienen uno, dos, o tres tipos de etiquetas, siendo difícil identificar un registro que presente una variedad de seis o siete tipos diferentes de etiquetas.

La siguiente categoría de etiquetas más observado en este gestor es el tipo de etiqueta refinadora. Estas etiquetas consisten en términos que por sí mismos no arrojan luz sobre la temática del documento, y que posiblemente hayan sido ingresados por los usuarios acompañando a otras etiquetas. Algunas de las etiquetas de este tipo más observadas en los registros son, por ejemplo, “Theory” o “Framework”. Por su parte, las etiquetas organizativas representan un espectro de etiquetas que incluyen cuestiones asociadas a dónde guardar el documento, para qué se lo utilizará, tareas asociadas al mismo, o incluso la fuente de la que ha sido tomado. Así, es posible observar en

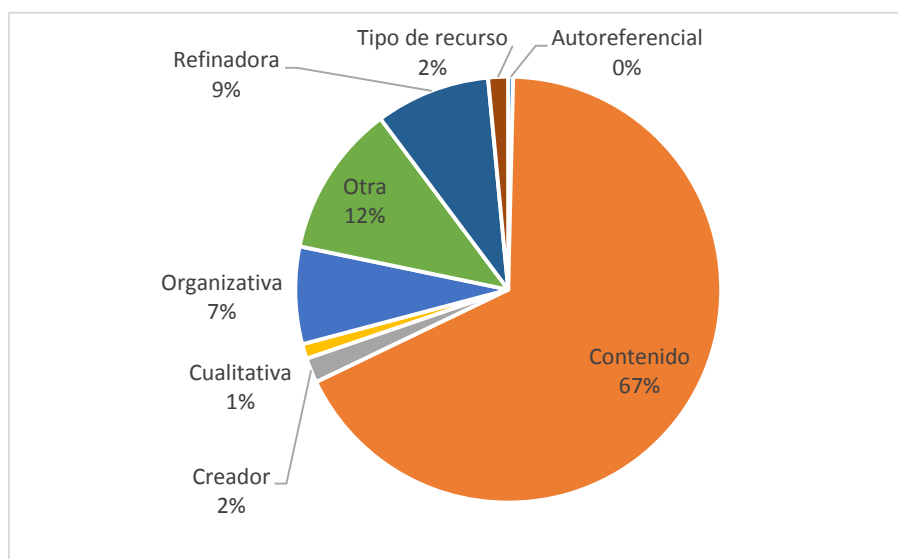


Gráfico 1. Tipo de etiquetas en CiteULike

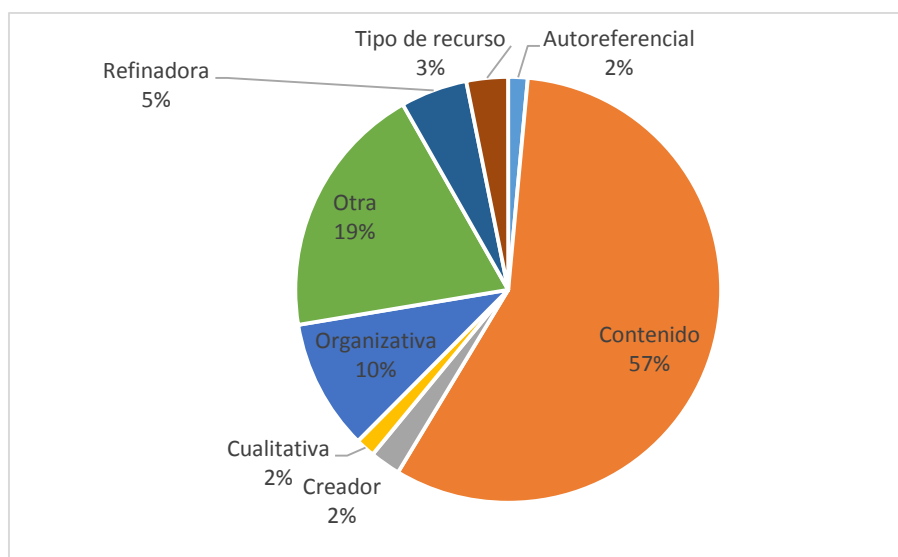


Gráfico 2. Tipo de etiquetas en Mendeley

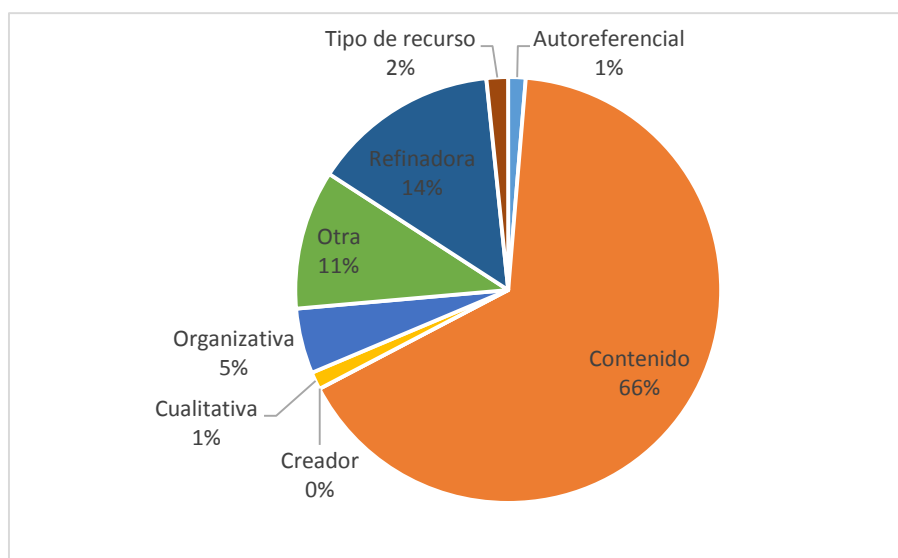


Gráfico 3. Tipo de etiquetas en Bibsonomy

diferentes registros la presencia de etiquetas tales como “FromCiteULike”, “Read”, o “Group-meeting”.

Las etiquetas menos observadas en este gestor incluyen las de Creador, Tipo de recurso, Cualitativa y Autoreferencial, en orden descendente. Las primeras (Creador) abarcan etiquetas que incluyen información sobre alguno de los creadores del documento, generalmente un apellido, o una combinación de apellido y nombre, o bien de apellido y otros datos no identificables (como por ejemplo la etiqueta “ferrari187401”, que presenta a “Ferrary, A.C.” como uno de sus autores). También fueron categorizadas como etiquetas de Creador aquellas que incluyen el nombre de la revista en la cual fue publicado originalmente el trabajo, como por ejemplo “Nature”. Las etiquetas de Tipo de recurso son aun más difíciles de encontrar, y abarcan palabras que dan indicios sobre el tipo de publicación, como por ejemplo “textbook”, o “journal”. Por su parte, las etiquetas cualitativas observadas incluyen voces que se intuyen como apreciaciones subjetivas de los recursos, tal como “general”, o “scary”. Se observan también en la muestra solamente dos etiquetas que se consideran como autoreferenciales. Estas son “cited_in_drm2009_by_me” y “Roberto”.

Por último, resulta interesante destacar que se observaron en la muestra una gran cantidad de etiquetas que no pudieron ser clasificadas en base al esquema original de Golder y Huberman, ya que no se ajustaban a ninguna de las categorías propuestas. Dichas etiquetas fueron clasificadas bajo el rótulo “Otra” e incluyen cuestiones variadas como: años de publicación, años no relacionables con el recurso, palabras de significado conocido pero que no pudieron ser asociadas a ninguna de las categorías establecidas, y palabras de significado desconocido que no pudieron ser asociadas al recurso, pero tampoco se tiene certeza de que puedan asociarse a un usuario en particular (es decir, que constituyan etiquetas autoreferenciales). Ejemplos de estas etiquetas son: “and”, “a5”, “mott”, “25-10-12”, “2006”, “gochebi”, entre otros.

Clasificación de etiquetas en Mendeley

Al igual que en el caso anterior, en el gestor Mendeley también se observa una importante predominancia de las etiquetas de contenido por sobre los demás tipos de etiquetas. Sin embargo, en este gestor el porcentaje de este tipo de etiquetas es ligeramente menor, dejando mayor espacio a la aparición de los demás tipos de etiquetas. Las etiquetas organizativas se observan en este gestor en mayor medida que en el caso anterior. Sin embargo, muchas de estas etiquetas incluyen también información que puede ser asociada a los contenidos de los recursos. Posiblemente, esto se deba a cuestiones estructurales del sitio web, que permite a los usuarios organizar sus colecciones en carpetas, por lo que muchas de las etiquetas asignadas a los recursos responden a los nombres de las carpetas en las que luego serán guardados por los usuarios. De esta forma, es posible encontrar a lo largo de la muestra una gran cantidad de etiquetas que comienzan con la palabra “Folder...”. Ejemplos de este tipo de etiquetas son: “Folder - Java”, o “Folder - Audience studies”. Sin embargo, muchas otras de estas etiquetas mantienen su carácter meramente organizativo, carente de significado para otros usuarios del sitio web, como por ejemplo “Folder - Methods - To look at in future”. Esto, además de las “clásicas” etiquetas Organizadoras, como por ejemplo “2read”.

Ya en mucha menor medida es posible observar etiquetas Refinadoras, de Tipo de recurso, de Creadores, Cualitativas o Autoreferenciales. En todos los casos, se trata de etiquetas muy similares a las ya estudiadas para el gestor anterior. Algunos ejemplos de todos estos tipos de etiquetas son (respectivamente): “Theoretical review”, “cookbook”, “A. K. Geim”, “best”, “Agatha”.

Posiblemente, uno de los aspectos más destacables del análisis de las etiquetas de este gestor es la gran cantidad de etiquetas encontradas que no pudieron ser asociadas a una categoría en particular del esquema original de clasificación. En este sentido, se observa también que además de los casos mencionados en el apartado anterior, relativos a fechas o palabras conocidas o no conocidas que no podían categorizarse, se presentan una gran cantidad de etiquetas compuestas por meros números o símbolos que a primera vista no parecieran tener ningún sentido o relación con el registro. Ejemplos de este tipo de etiquetas son: “+”, “5414703”, “*****”, “4/203”, “?”, “[2]”, “061110_1”, solo por destacar algunos ejemplos de las 92 etiquetas que fueron clasificadas bajo la categoría “Otra”.

Clasificación de etiquetas en Bibsonomy

La distribución de etiquetas observada en este gestor bibliográfico presenta un alto grado de similitud con la distribución presente en el primer gestor analizado, CiteULike. Si bien el primer gestor contiene una cantidad mucho mayor de etiquetas, de casi el doble que Bibsonomy, es posible verificar que en valores porcentuales las etiquetas presentan una distribución muy similar. En este sentido, se observa que las etiquetas de contenido representan un 66% del total de la muestra (solo un punto menos que en CiteULike). Posiblemente, la principal diferencia se observa en la presencia de una gran cantidad de etiquetas refinadoras, es decir, de etiquetas que no presentan “poder de descripción” por sí mismas. Este tipo de etiquetas abarca el 14% de la muestra, dejando solo un 20% para los otros cinco tipos de etiquetas, más las etiquetas clasificadas como “Otras”. Además, se observa la presencia de etiquetas refinadoras mucho más específicas en relación a los contenidos del registro, como por ejemplo “mining”, “properties”, o “reduced”. Este tipo de etiquetas permite intuir una clara intención por parte de los usuarios de conformar etiquetas multitérminos que, de alguna forma, terminaron en etiquetas de un solo término al cual le falta su acompañante para formar la etiqueta completa (“data mining”, “graphene properties”, y “oxide reduced” en base a los ejemplos mencionados). Posiblemente, esta característica esté relacionada con el hecho de que el sitio web no admite etiquetas compuestas por palabras separadas por espacios. Esto pudo haber llevado a que muchos usuarios ingresaran etiquetas múltiples, cuando en realidad querían ingresar una única etiqueta, compuesta de distintas palabras separadas por espacios.

Para todos los demás tipos de etiquetas, como las Organizativas, las de Tipo de recurso, las Cualitativas y las Autoreferenciales, se observa una similitud con las etiquetas ya estudiadas en los dos gestores anteriores. Algunos ejemplos de estas etiquetas son (respectivamente): “toread”, “publication”, “wanted”, y “myown”.

A diferencia de los dos gestores anteriores, en Bibsonomy no se pueden identificar etiquetas relacionadas con el Creador del recurso.

Otro aspecto que merece ser destacado es que este es el único gestor analizado que presenta registros en los que no se observa una etiqueta de contenido. Se trata de dos registros que presentan una única etiqueta, de tipo organizativa: “imported”.

Por último, cabe también destacar que este gestor es el que presenta la mayor cantidad de etiquetas adecuables al esquema original de clasificación. Las etiquetas clasificadas como “Otras” constituyen solo un 11% del total, lo que representa casi la mitad de las etiquetas de este tipo observadas en Mendeley, y 1 punto menos que las etiquetas de este tipo observadas en CiteULike. Al igual que en CiteULike (y a diferencia de Mendeley) las etiquetas de este tipo se componen por fechas que pueden o no estar asociadas a los recursos y, principalmente, por palabras desconocidas que no pueden relacionarse directamente con los recursos, tales como “dblp”, o “2010”. Se observa

CiteULike		Mendeley		Bibsonomy	
Tipo de etiqueta	Cantidad	Tipo de etiqueta	Cantidad	Tipo de etiqueta	Cantidad
Altruistas	388	Altruistas	271	Altruistas	229
Egoistas/Amiguistas	141	Egoistas/Amiguistas	203	Egoistas/Amiguistas	74
Populistas	0	Populistas	0	Populistas	0
Total de etiquetas	529	Total de etiquetas	474	Total de etiquetas	303

Tabla 2: Tipos de etiquetas por gestor. Fuente: elaboración propia

una única etiqueta compuesta por símbolos, la etiqueta “\&”, que al igual que en el caso de las etiquetas refinadoras, pareciera haber sido ideada para formar parte de una etiqueta mayor, pero fue ingresada de manera independiente por error.

Clasificación según el esquema de Cañada

Más allá del análisis realizado en la sección anterior, en el cual se pudo conocer sobre qué versan las etiquetas presentes en los gestores bibliográficos, resulta interesante observar si esas etiquetas pueden considerarse útiles para la comunidad general de usuarios de esos sitios web. A prima facie podría pensarse que aquellas etiquetas que hacen referencia a los contenidos abordados en el recurso podrían ser las más útiles para la comunidad de usuarios. Sin embargo, si bien esto es así en la mayoría de los casos, existen muchos otros casos en los que los contenidos son reflejados en las etiquetas en formas peculiares, lo cual dificulta su utilización por parte de otros usuarios. A la inversa, algunas etiquetas que se ubican en los grupos de “Etiquetas refinadoras” o bien de “Etiquetas de creador”, pueden contener información útil para otros usuarios del sitio web.

Para echar luz sobre estas cuestiones se aplica la segunda parte de esta sección de la matriz de análisis, en la cual se clasifican las etiquetas en base a una variación del esquema de clasificación de Cañada, para determinar el grado de utilidad de las etiquetas recogidas en la muestra.

Tal como puede observarse en la Tabla 2 ninguno de los tres gestores presentan en su muestra etiquetas que puedan considerarse como “populistas”, es decir, que no representen verdaderamente el contenido de los recursos y hayan sido incluidas para aumentar el nivel de visualización del mismo. Posiblemente, esto esté relacionado al hecho de que los gestores bibliográficos se caracterizan por ser recursos de uso principalmente académico, en los que priman intereses diferentes a los que normalmente suelen promocionarse en las etiquetas engañosas.

Así mismo, es posible apreciar que en los tres gestores bibliográficos se observa una mayor cantidad de etiquetas altruistas, que de etiquetas egoístas/amiguistas, aunque en diferentes cantidades.

El Gráfico 4 permite apreciar una comparación a nivel de porcentajes, más allá de los números absolutos, del tipo de etiquetas presentes en cada gestor.

Es posible observar que de los tres gestores bibliográficos Mendeley aparece como el gestor en el que se observan mayor cantidad de etiquetas egoístas/amiguistas,

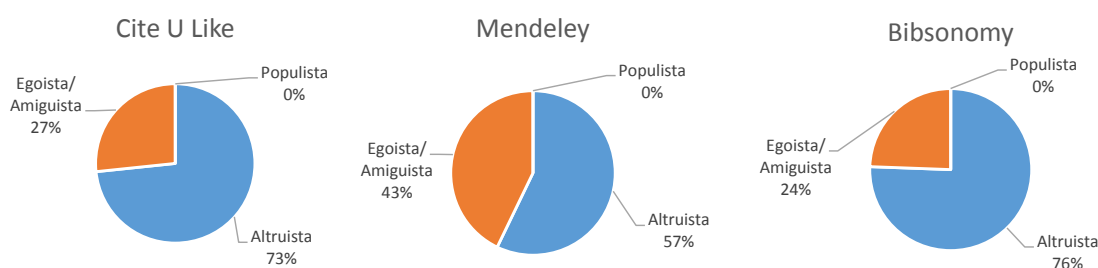


Gráfico 4. Comparación del tipo de etiquetas

mientras que CiteULike y Bibsonomy comparten un porcentaje muy similar, en donde estas etiquetas representan aproximadamente solo un cuarto de la muestra.

Las etiquetas consideradas altruistas se componen principalmente de etiquetas de contenido, que ofrecen información explícita y comprensible por todos los usuarios respecto de las temáticas abordadas por los recursos en cuestión. Estas etiquetas pueden estar compuestas de un solo término, o bien de más de un término, unidos de diferentes formas (mediante el uso de un guión, con espacios, sin espacios, etc). Además, se incluyen dentro de este tipo de etiquetas diferentes idiomas y alfabetos, y hasta variaciones ortográficas (por ejemplo en los usos de “blog” y “blogg”). Ejemplos de este tipo de etiquetas son: “gender”, “biomedicine”, o “electronic”.

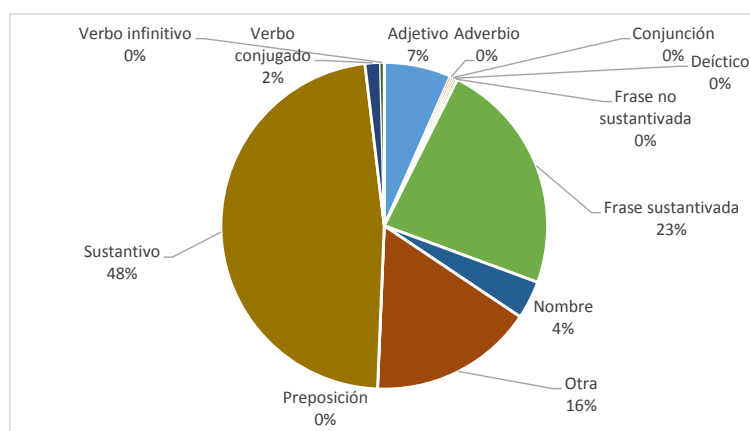
Por otro lado, se incluyen también dentro de las etiquetas altruistas algunas etiquetas que no apuntan directamente al contenido, pero que claramente fueron asignadas con la función de clarificar de qué trata el recurso. Es el caso de las etiquetas que en la sección anterior fueron identificadas como etiquetas refinadoras, que pueden ayudar a la comunidad de usuarios a interpretar el contenido del recurso en cuestión. Ejemplos de este tipo de etiquetas son las etiquetas “research”, “application” y “properties”, que se presentan en interacción directa con las etiquetas mencionadas en el párrafo anterior.

Finalmente, se incluyen también en esta categoría etiquetas de creador, e incluso algunas etiquetas clasificadas bajo la categoría “Otras”, que proveen información sobre el autor del recurso o el año de su publicación. Ejemplos de estas etiquetas son: “Lihong Wang”, o “2010”.

Por su parte, las etiquetas egoístas/amiguistas incluyen principalmente etiquetas que fueron identificadas en la sección anterior como etiquetas organizativas, cualitativas y autoreferenciales. Ejemplos de estos tipos de etiquetas son: “Hard copy in Irv’s office”, “myown”, o “scary”. También aquí se incluyen muchas de las etiquetas que en la sección anterior fueron identificadas como “Otras”, tales como aquellas basadas en símbolos, o palabras incomprensibles.

Por último, se incluyen también dentro de esta categoría muchas etiquetas que, si bien se interpreta que tienen como finalidad última aportar pistas sobre el contenido del recurso, la modalidad en la que lo hacen da cuenta de etiquetas que fueron pensadas para ser recuperadas por el propio usuario o su grupo de pertenencia, y no por la comunidad en general. Ejemplos de este tipo de etiquetas son: “rgo” (por Reduced Graphene Oxide Nanowires), “anth” (por Anthropology), o “bioMOF” (por Metal-organic frameworks in biomedicine). Se entiende que la escritura particular de estas

Gráfico 5. Construcción lingüística en CiteULike



etiquetas va más allá de la simple unión de palabras o las variaciones ortográficas, llegando a ser deformaciones personales o profesionales de las mismas, y por tanto se las incluye como etiquetas “egoístas/amiguistas” que no apuntan a aumentar la capacidad de recuperación de los demás usuarios del sitio web.

Elaboración de las etiquetas

Elaboración lingüística de las etiquetas

En los apartados anteriores hemos podido conocer a qué se refieren las etiquetas creadas por los usuarios (si al contenido, a aspectos organizativos, etc.) y también si las mismas podrían ser de utilidad para todos los usuarios del sitio web, o solo para quién las creó y su grupo de pertenencia. Sin embargo, resulta también interesante conocer cómo son construidas las etiquetas. En este sentido, es sabido que las etiquetas que tienen mayor poder de descripción son aquellas que se corresponden con sustantivos o frases nominales sustantivadas. Mientras que otras construcciones lingüísticas, como los adjetivos, los artículos, las preposiciones, o los adverbios, entre otros, no suelen utilizarse de forma solitaria para la construcción de voces temáticas, por carecer de valor intrínseco de descripción.

A partir de la aplicación de la segunda sección de la matriz de análisis, es posible conocer cuáles son las formas lingüísticas utilizadas para construir las etiquetas que formaron parte de la muestra, y también cuáles son las más utilizadas. Tal como se puede observar en la Tabla 3, los resultados del análisis demuestran que las construcciones más utilizadas son aquellas que se corresponden con el uso de sustantivos y frases sustantivadas. En el caso de CiteULike, el uso de sustantivos supera ampliamente al resto de las construcciones lingüísticas, siendo aun más del doble que las frases sustantivadas. En el caso de Bibsonomy esta diferencia se acrecienta llegando a representar los sustantivos más de cuatro veces la cantidad de frases sustantivadas. Finalmente, en el caso de Mendeley, se observa una preponderancia algo mayor al uso de frases sustantivadas, que supera en un pequeño número al uso de sustantivos simples.

Elaboración lingüística de las etiquetas en CiteULike

Tal como se puede observar en el Gráfico 5, casi la mitad de las etiquetas de la muestra de CiteULike (48%) se pueden asociar al uso de sustantivos simples.

Este tipo de etiquetas son aquellas que se componen de un único término a partir del cual se designa de forma directa un elemento, una idea, o un ser vivo. Se incluyen dentro de esta categoría sustantivos comunes, individuales, colectivos, propios, concretos y

CiteULike		Mendeley		Bibsonomy	
Construcción	Cantidad	Construcción	Cantidad	Construcción	Cantidad
Adjetivo	35	Adjetivo	13	Adjetivo	31
Adverbio	1	Adverbio	0	Adverbio	0
Conjunción	1	Conjunción	1	Conjunción	1
Deíctico	1	Deíctico	0	Deíctico	4
Frase no sustantivada	1	Frase no sustantivada	1	Frase no sustantivada	0
Frase sustantivada	123	Frase sustantivada	154	Frase sustantivada	40
Nombre	20	Nombre	14	Nombre	6
Otra	86	Otra	133	Otra	38
Preposición	0	Preposición	0	Preposición	2
Sustantivo	251	Sustantivo	149	Sustantivo	153
Verbo conjugado	8	Verbo conjugado	7	Verbo conjugado	21
Verbo infinitivo	2	Verbo infinitivo	2	Verbo infinitivo	7
Total general	529	Total general	474	Total general	303

Tabla 3: Construcción de etiquetas por gestor. Fuente: elaboración propia

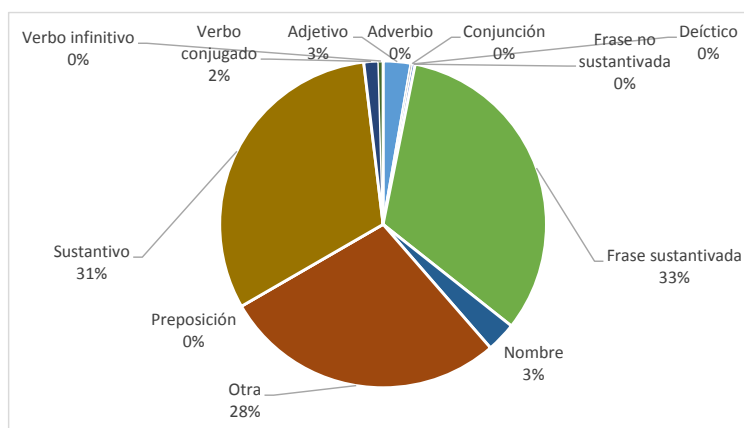
abstractos. Ejemplos de estos tipos de etiquetas son: “impact”, “histogram”, “systems”, “power”, o “gender”, por nombrar algunos ejemplos.

Otra importante cantidad de etiquetas se componen de sintagmas que incluyen un sustantivo, acompañado de otro elemento, generalmente un adjetivo u otro sustantivo coordinado. Dichas construcciones fueron categorizadas bajo el nombre de “frases sustantivadas” y representan el 23% del total de la muestra. Algunos ejemplos de estas etiquetas son: “sex-role”, “rise-of-graphene”, o “text-mining”. En este punto resulta interesante señalar que muchas de las construcciones que conforman los sintagmas se presentan en las etiquetas como palabras unidas, sin ningún tipo de separación. En el presente análisis se ubica a aquellos términos cuyo uso natural implica el uso de una palabra compuesta bajo la categoría “sustantivo”, mientras que las uniones “no convencionales” de palabras se consideran como “frases sustantivadas”. En este sentido la etiqueta “biomedicine” (unión de las palabras “bio” y “medicine”), cuyo uso extendido valida la unión de ambas palabras para formar un único término de significado propio, se considera como un sustantivo. Mientras que las etiquetas “medicalanthropology” o “savageminds”, cuyo uso conjunto no se encuentra extendido, sino que se tratan de uniones ad-hoc generadas por el usuario, se consideran frases sustantivadas.

Otra salvedad que resulta interesante destacar es el caso de las frases no sustantivadas. Es decir, frases cuya construcción no se encuentra centrada en el sustantivo sino en un sujeto o una acción. En el caso de CiteULike se observa un ejemplo de dicha construcción en la etiqueta “cited_in_drm2009_by_me”.

Ya en un porcentaje mucho menor, cercano al 7 % del total de la muestra, es posible observar la presencia de etiquetas compuestas por adjetivos utilizados en forma simple, es decir, sin el acompañamiento de un sustantivo. Como se mencionó en párrafos anteriores este tipo de etiquetas resultan poco útiles para la recuperación de información, ya que carecen de poder de descripción por sí mismas. Ejemplos de estos tipos de etiquetas son: “new”, “scientific”, o “general”.

Gráfico 6 Construcción lingüística en Mendeley



Aún más controvertido que el caso de los adjetivos utilizados en soledad, es el caso de los deícticos. Como es sabido, los términos que componen una etiqueta de contenido deben constituir unidades autónomas y completas. Por tal motivo, no suelen incluirse en los vocabularios temáticos etiquetas que hagan referencia a otros términos no incluidos en la misma voz. En el caso de CiteULike solo se ha encontrado una etiqueta con estas características: “fixme”.

También ha sido posible observar en la muestra un porcentaje muy pequeño, cercano al 2 % del total, de etiquetas generadas en base al uso de verbos. Si bien el uso de estas construcciones lingüísticas es normalmente aceptado en los vocabularios temáticos, en líneas generales se prefiere que los verbos se utilicen en su forma nominal. Más allá de eso, ha sido posible detectar en la muestra de CiteULike el uso de verbos en tiempo infinitivo tales como “contact” y aun en mayor medida verbos conjugados como “modeling”, o “selected”.

Así mismo, también ha sido posible advertir la presencia de una etiqueta compuesta por una conjunción. Por supuesto, esto resulta altamente llamativo ya que por lo general las conjunciones se utilizan para conectar dos sustantivos. De más está decir que el uso de una conjunción por sí misma no tiene ningún tipo de utilidad para la recuperación de la información. Se trata en este caso de un único registro en el cual es posible observar la etiqueta “and”.

Además de todas las formas lingüísticas mencionadas, también es posible observar en la muestra la presencia de nombres propios, tanto personales como “novoselov” o “Williams”, como así también de sitios web o publicaciones, tales como “citeulike”, “flickr”, o “Nature”.

Finalmente, al igual que en los análisis realizados en la sección anterior, fue necesario también en esta sección incluir una categoría denominada “Otra” para identificar todas aquellas etiquetas que no pudieron ser categorizadas dentro de las construcciones lingüísticas propuestas. Se trata principalmente de etiquetas compuestas por números o por palabras cuyo significado resulta desconocido, o palabras que presentan serias deformaciones, como es el caso de las etiquetas “2010”, “cnt”, “ngs”, “t-retrieval”, “bionlp”, o “_result_3”, por mencionar algunos ejemplos.

Elaboración lingüística de las etiquetas en Mendeley

En el Gráfico 6 es posible apreciar como en este gestor la totalidad de etiquetas que componen la muestra se encuentra dividida en una forma mucho más homogénea

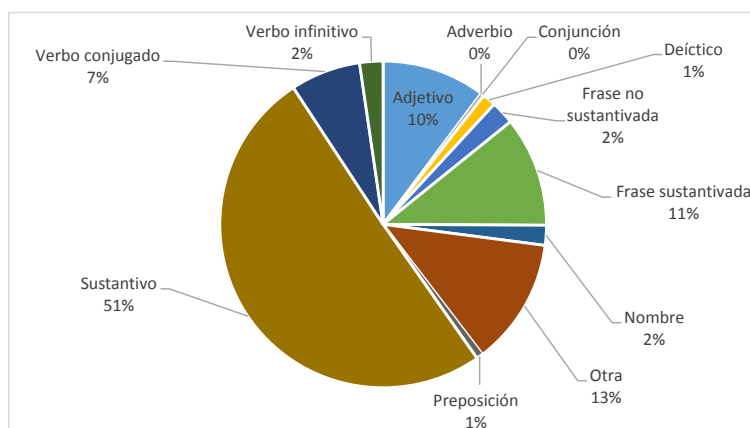


Gráfico 7 Construcción lingüística en Bibsonomy

entre tres categorías sobresalientes: las frases sustantivadas (33%), los sustantivos (31%) y las etiquetas categorizadas como “Otras”, es decir, que no corresponden a ninguna de las construcciones lingüísticas analizadas (28 %).

La principal diferencia entre este gestor y el gestor CiteULike analizado en el apartado anterior se presenta en el hecho de que aquí puede observarse una presencia mucho mayor de etiquetas compuestas por más de un término. Se trata de los sintagmas con núcleo sustantivo, como es el caso de las etiquetas “Agenda Setting”, “applications graphene”, o “Biomedical technology”, por poner algunos ejemplos.

Ya en mucha menor medida, con promedios de 3 % cada una, es posible observar las categorías de etiquetas basadas en nombres propios y en adjetivos. Aún en menor medida, cercana al 2 %, se puede observar también la presencia de verbos conjugados y verbos infinitivos. Algunos ejemplos de estos tipos de etiquetas en este gestor son: “Bert Bates”, “Twitter”, “Bio”, “antiguos”, “downloaded”, o “imaging”.

También en este gestor, al igual que en el anterior, es posible observar la presencia de una etiqueta compuesta por una conjunción. Se trata de hecho de la misma etiqueta “and”. Así mismo, también es posible observar la presencia de una etiqueta cuya composición constituye una frase no sustantivada: la etiqueta “Hard copy in Irv’s office”.

Por último, uno de los aspectos que resulta de mayor interés para destacar es la gran presencia de etiquetas que fueron designadas bajo la categoría “Otras”. Estas etiquetas constituyen casi un tercio del total de la muestra, y no solo se encuentran compuestas por números o palabras de significado desconocido, sino por etiquetas que directamente no contienen palabras. Tal es el caso de etiquetas como “***”, “?”, “#2”, o “-”, entre otras. Por supuesto, estas etiquetas no representan absolutamente ningún tipo de ayuda para la recuperación de información.

Elaboración lingüística de las etiquetas en Bibsonomy

En el Gráfico 7 es posible apreciar el uso de construcciones lingüísticas para este gestor en particular.

Tal como se puede observar, la mitad de la muestra se compone de etiquetas que se corresponden con la forma de sustantivos. Sin embargo, resulta interesante destacar que, a diferencia de los casos anteriores, se observa una repartición mucho más uniforme entre los demás tipos de construcciones lingüísticas analizadas. Se observa una fuerte presencia de etiquetas construidas en forma de frases sustantivadas, pero

también de adjetivos utilizados en soledad, y de formas verbales conjugadas y en infinitivo. Ejemplos de algunas de estas etiquetas en este gestor son: “socialnetworks”, “enterpriseapplication”, “efficient”, “free”, “engineering”, o “imported”.

A diferencia de los casos anteriores, no se observa en este gestor la presencia de frases no sustantivadas. Sin embargo, es posible apreciar una presencia más fuerte de deícticos que en los casos anteriores. La muestra presenta un total de 4 etiquetas construidas bajo esta forma correspondientes a: “myown” y “mykopie”.

También en este gestor es posible observar la presencia de la etiqueta “and”, constituida por una mera conjunción, en uno de los registros. Así mismo, se ha identificado en este gestor una nueva forma verbal que no se había hecho presente en los anteriores casos. Se trata de etiquetas compuestas por preposiciones que, al igual que las conjunciones, no tienen ningún poder de descripción de los contenidos, ni representan una ayuda para la futura recuperación de la información. Se trata de la etiqueta “diss”, presente en dos registros.

La presencia de nombres propios resulta en este caso menor que en los casos anteriores, cercana al 1 % del total de la muestra. Ejemplos de estas etiquetas son: “Safari” (por el nombre del explorador de Apple), o “Ferghana”.

Finalmente, se observa una pequeña porción de la muestra, cercana al 13%, de etiquetas designadas bajo la categoría “Otras”. Si bien es posible observar la presencia de alguna etiqueta compuesta por símbolos, como el caso de la etiqueta “\&”, la mayoría de estas etiquetas se corresponden con números o palabras desconocidas o deformadas por gustos personales. Ejemplos de este tipo de etiquetas son “etnici”, “cntfets”, o “jrr”.

Forma y escritura de las etiquetas

Además de conocer cuál es la estructura lingüística que subyace en las etiquetas, resulta también interesante conocer otros aspectos relacionados a su construcción, particularmente aquellos ligados a la forma y la escritura de las etiquetas. Específicamente, cuestiones tales como el uso de la forma singular y plural, el uso de mayúsculas y minúsculas, y la forma en que se conjugan las palabras en las etiquetas compuestas por más de un término. Esto deriva en un aspecto más profundo, relacionado con la cantidad de veces que una misma palabra es ingresada con formas o escrituras diferentes, generando múltiples etiquetas para lo que es, esencialmente, el mismo término. Ello, sin siquiera entrar en terreno de un análisis semántico en el que se identifique el uso de sinónimos y cuasi-sinónimos para representar un mismo concepto, sino de un análisis meramente sintáctico en base al uso de las mismas palabras.

Al aplicar la segunda parte de esta sección de la matriz de análisis es posible reconocer cuáles son las variaciones más observadas en cada uno de los gestores bibliográficos estudiados, y cuál es el grado de re-utilización de las etiquetas ya creadas por otros usuarios (en contraposición a la creación de nuevas etiquetas con las mismas palabras).

Tal como puede observarse en la Tabla 4, el gestor bibliográfico Mendeley es el que presenta la mayor cantidad de variaciones observadas.

En este punto, resulta importante destacar que, tal como se indicó en el párrafo anterior, estas variaciones fueron observadas específicamente en los casos de etiquetas compuestas exactamente por las mismas palabras. Esto implica que el presente análisis no considera como variaciones casos como:

Tipo de variación	Cantidad observada		
	CiteULike	Mendeley	Bibsonomy
Forma plural/singular	12	4	3
Variación del idioma	4	7	2
Palabra cortada	6	1	1
Unión por guión	5	5	
Variación de ortografía	2	3	
Uso de paréntesis		1	1
Uso del punto		2	
Uso de la barra			1
Uso de la coma			3
Uso de corchetes			1
Forma mayúscula/minúscula		22	
Unión por espacio		1	
Unión por guión y mayúscula/minúscula		2	
Variación del idioma y mayúscula/minúscula		1	
Forma plural/singular y mayúscula/minúscula		2	

Tabla 4. Variaciones observadas

1. Etiquetas compuestas por un término, y etiquetas compuestas por ese mismo término en conjunto con otros. Por ejemplo:

Antropología
Antropología cultural
Antropología política

2. Etiquetas compuestas por términos que implican distintas flexiones de una misma raíz. Por ejemplo:

Crime
Criminality
Criminalization

3. Etiquetas compuestas por dos o más términos, y etiquetas compuestas por esos mismos términos pero utilizados de manera independiente. Por ejemplo:

Breast cancer
Breast
Cancer

4. Etiquetas que implican sinónimos o cuasi-sinónimos. Por ejemplo:

Biomedicine
Pharmacy

Forma y escritura de las etiquetas en CiteULike

Tal como puede apreciarse en la Tabla 4, la variación más observada en la muestra de etiquetas de este gestor es el uso simultáneo de la forma plural y la forma singular. En 12 ocasiones se detectaron etiquetas escritas tanto en forma plural como en forma singular. Esto implica que el gestor posee 24 etiquetas para representar exactamente el mismo término cuando, muy posiblemente, solo debería tener 12. Algunos de los 12 casos de variación detectados son:

Forma singular	Forma plural
communication	communications
composite	composites
dissertation	dissertations
image	images
network	networks

Tal como se puede observar, algunos de estos términos podrían llegar a representar conceptos diferentes, haciendo necesario el uso de la forma plural y singular en simultáneo. Es el caso de las etiquetas “communication” y “communications”, en donde se podría llegar a pensar que la forma en singular es utilizada para expresar un proceso, mientras que la forma plural es utilizada para expresar los productos de ese proceso. Por supuesto, aseverar esta situación implicaría un análisis que escapa a los alcances de este trabajo. En cualquier caso, para la mayoría de las variaciones observadas en el uso de plural y singular, no se distinguen motivos válidos para el uso conjunto de ambas formas.

Otra forma común de variación, que genera duplicación y hasta triplicación de etiquetas, es el uso de palabras cortadas. En 6 ocasiones diferentes se pudo observar la existencia de etiquetas compuestas por una palabra, junto con etiquetas compuestas por la misma palabra, pero de la cual solo componían la etiqueta las primeras letras de la palabra. Ejemplos de este tipo de variación son:

Palabra completa	Palabra cortada
Experiment	Exp
Anthropology	Anth
	Anthro

Así mismo, dado que el gestor no permite ingresar espacios entre los términos de una etiqueta, se han observado casos en los que se han utilizado distintos tipos de guiones para unir las palabras. Algunos ejemplos de estas variaciones son:

Forma 1	Forma 2
new_media	new-media
socialmedia	social-media
textmining	text-mining

Por su parte, las variaciones idiomáticas podrían verse más justificadas debido al uso universal de este tipo de recursos. Sin embargo, no deja de llamar la atención la

existencia de etiquetas de traducción textual a otros idiomas que pueden encontrarse en la muestra. Algunos ejemplos son:

Idioma 1	Idioma 2
familien	families
minería	mining
social_medie	social_media

Más extrañas resultan las variaciones ortográficas, es decir, palabras escritas en un mismo idioma, pero con una variación en la ortografía utilizada. En la muestra de este gestor se detectan solo dos casos, que se copian a continuación:

Forma 1	Forma 2
blogg	blog
pluralization	pluralisation

Como contrapunto de las variaciones observadas, resulta interesante señalar que solo 43 etiquetas fueron utilizadas en más de una oportunidad, constituyendo un total de 158 usos de etiquetas, de los 529 estudiados en la muestra. Las 371 etiquetas restantes solo fueron utilizadas en una única oportunidad. Esto implica que en menos del 30% de las ocasiones en las que los usuarios debieron ingresar una etiqueta, eligieron utilizar una etiqueta que ya existía en el sistema. En cambio, en más del 70% de los casos los usuarios ingresaron nuevas etiquetas, diferentes a las ya existentes, aun cuando las variaciones en las etiquetas eran mínimas (como las estudiadas en los párrafos anteriores).

De todas las etiquetas que fueron reutilizadas, la mayor cantidad de re-usos observada es de 10 usos. Se observan solo 4 etiquetas que fueron utilizadas 10 veces: “anthropology”, “graphene”, “java” y “media”. Se reproduce a continuación una tabla de todas las etiquetas que han sido utilizadas en más de una oportunidad en la muestra (Tabla 5).

Forma y escritura de las etiquetas en Mendeley

Tal como se puede observar en la Tabla 4 el gestor bibliográfico Mendeley es el que presenta la mayor cantidad de variaciones observadas. De todos los tipos de variaciones, la que se hace presente con mayor fuerza es aquella en la que varía el uso de las mayúsculas y las minúsculas. Esta variación que parece tan frecuente en este gestor, paradójicamente no figura entre las variaciones de los otros dos gestores estudiados. El motivo por el que esto sucede es de hecho muy simple: Mendeley es el único de los tres gestores estudiados que se comporta de manera “case-sensitive”, es decir, que distingue el uso de mayúsculas y minúsculas, y lo mantiene al momento de generación de las etiquetas. Tanto CiteULike como Bibsonomy generan todas las etiquetas en letras minúsculas, evitando así la generación de etiquetas duplicadas por uso simultáneo de letras mayúsculas y minúsculas.

El uso de estas dos formas de escritura genera en Mendeley hasta cuatro tipos de etiquetas diferentes: etiquetas escritas enteramente en minúscula, etiquetas escritas enteramente en mayúscula, etiquetas escritas con la primera letra en mayúscula y el resto en minúscula, y en los casos de etiquetas multi-término, etiquetas que utilizan la mayúscula para la primera letra de cada palabra y la minúscula para el resto de las letras.

Etiqueta	Usos	Etiqueta	Usos	Etiqueta	Usos	Etiqueta	Usos
anthropology	10	medical	4	anth	2	integration	2
graphene	10	biomedical	3	biology	2	materials	2
java	10	digital	3	book	2	med	2
media	10	flexible-electronics	3	compiler	2	medicine	2
review	9	history	3	composite	2	ngs	2
biomedicine	8	language	3	elumen	2	textbook	2
theory	7	macroelectronics	3	experiment	2	textmining	2
social	6	nature	3	featherweight_java	2	text-mining	2
mass	5	semantics	3	gender	2	thesis	2
bioinformatics	4	software	3	generics	2	web	2
communication	4	2006	2	impact	2		

Tabla 5: Etiquetas más utilizadas en CiteULike. Fuente: elaboración propia

Algunos ejemplos de estas variantes son:

Forma 1	Forma 2	Forma 3
ANTHROPOLOGY	Anthropology	anthropology
COMMUNICATION	Communication	communication
Big Data	Big data	big data

El igual que en el gestor CiteULike, también en la muestra de este gestor es posible observar variaciones idiomáticas, con la particularidad de que en este caso las variaciones llegan a presentarse incluso en el alfabeto utilizado. Algunos ejemplos de estas variaciones son:

Idioma 1	Idioma 2
プログラミング	programming
программирование	programming
Ява	java

Así mismo, también pueden encontrarse variaciones en la ortografía utilizada para escribir una palabra en el mismo idioma. Algunos ejemplos son:

Forma 1	Forma 2
applicationsgraphene	aplicationsgraphene
biomedicalization	biomedicalisation
devices	devises

Otra forma variante encontrada en la muestra es el uso simultáneo de la forma plural y la forma singular. Algunos ejemplos de esta variación son:

Forma singular	Forma plural
application	applications
book	books
computer	computers

Respecto a la unión de palabras para formar etiquetas de términos compuestos, se observan formas variantes en las que se unen palabras simplemente colocando una seguida de la otra, o bien mediante el uso de guiones o el uso del espacio. Ejemplos de estas variantes son:

Forma 1	Forma 2
agenda setting	agenda-setting
Communication Social aspects	Communication - Social aspects
Downes2010	Downes_2010
Media Malaise	Mediamalaise

En este gestor se observa además la aparición de otras variantes no encontradas en el análisis de CiteULike, como el uso de signos de puntuación. En este caso, se observa el uso del paréntesis y del punto. Los casos encontrados se transcriben a continuación:

Forma 1	Forma 2
(computer	computer
anthropology.	anthropology
Anarchism.	Anarchism

Es plausible pensar que en algunos casos (sobre todo en el primero) la inclusión de estos signos puede deberse a algún tipo de error en la creación de etiquetas por parte del usuario, que posiblemente pretendía ingresar muchas palabras en una misma etiqueta, pero terminó generando etiquetas independientes. Una pista de ello es la existencia en el mismo registro de la etiqueta “language)”, cuyo paréntesis final viene a completar una etiqueta inexistente que quedaría formada como “(computer language)”.

Finalmente, resulta llamativo observar la presencia de “variaciones múltiples”, cosa que no es posible identificar en los otros dos gestores analizados en este trabajo. Se trata de etiquetas que varían unas de otras en más de un aspecto, generándose formas triplicadas y hasta cuadruplicadas de ingresar una misma palabra o conjunto de palabras. Algunas de estas variaciones observadas son:

Etiqueta	Usos	Etiqueta	Usos	Etiqueta	Usos	Etiqueta	Usos
anthropology	10	culture	3	3 todo	2	biophotonics	2
Anthropology	6	Folder - java	3	3antropologia	2	Book chapter	2
java	6	1	2	Agatha	2	carbon	2
(computer	4	870	2	agenda setting	2	Carbon: chemistry	2
Java	4	(guión medio)	2	anthropological theory	2	checked	2
Antropologia cultural	4	Agenda Setting	2	Anthropology	2	computer	2
book	4	Antropologia	2	anthropology.	2	computers	2
Folder - 10-Java	4	Applications	2	Antropologia del parentiu	2	graphene	2
Antropologia de les migracions	3	Book	2	Antropologia política	2	Lihong Wang	2
application	3	COMMUNICATION	2	Antropologia social	2	mass media	2
applications	3	Graphene	2	Au	2		
biomedicine	3	1SEA	2	Big data	2		

Tabla 6: Etiquetas más utilizadas en Mendeley. Fuente: elaboración propia

Variación	Forma 1	Forma 2
Unión por guión y mayúscula/minúscula	COMMUNICATION -- Social aspects	Communication Social aspects
Unión por guión y mayúscula/minúscula	Agenda-setting	agenda setting
Variación del idioma y mayúscula/minúscula	Antropologia	anthropology
Forma plural/singular y mayúscula/minúscula	Applications	application

Respecto del uso de etiquetas ya ingresadas por otros usuarios, puede observarse en el siguiente gráfico que el número de etiquetas utilizadas en más de una oportunidad decrece en comparación con los porcentajes observados para el gestor CiteULike, siendo menor al 26% del total de la muestra.

Resulta interesante destacar que el número total de etiquetas que fueron utilizadas en más de una oportunidad es incluso superior al número encontrado en CiteULike, siendo un total de 46 etiquetas. Esto resulta llamativo si se considera que la muestra total de Mendeley es bastante menor que la muestra total de etiquetas de CiteULike. Esta diferencia que aparece entre un número mayor de etiquetas utilizadas en más de

una oportunidad, pero un número menor de porcentaje de reutilización de etiquetas se debe a que en Mendeley, las etiquetas reutilizadas son más, pero son utilizadas en menor medida. En este sentido solo una etiqueta de toda la muestra, “anthropology”, es utilizada 10 veces (la cantidad máxima observada), a diferencia de CiteULike en donde se podía observar 4 etiquetas con este nivel de uso. Se reproduce la Tabla 6 con todas las etiquetas que han sido utilizadas en más de una oportunidad en la muestra.

Forma y escritura de las etiquetas en Bibsonomy

La muestra del gestor bibliográfico Bibsonomy es, de las tres estudiadas, la que menos variaciones presenta. Tal como se observa en la Tabla 4 una de las principales variaciones observadas es el uso simultáneo de la forma plural y singular. Aun así, solo se identifican tres casos en los que se hace presente esta variación. Ellos son:

Forma singular	Forma plural
economic	economics
computer	computers
citation	citations

Otra variación que también se encontró es aquella en la que se presenta una misma etiqueta, en diferentes idiomas. Los dos casos encontrados son:

Idioma 1	Idioma 2
algorithm	algorithmen
partecipazione	partizipation

Resulta llamativo que, a diferencia de los otros dos gestores estudiados, las principales variaciones se observan en el uso adicional de algún signo de puntuación o símbolo, tales como los paréntesis, la barra, la coma o los corchetes. Estas variaciones son:

Variación	Forma 1	Forma 2
Uso de paréntesis	(rgo)	rgo
Uso de la barra	history/	history
Uso de corchetes	{information}	information
Uso de la coma	history,	history
	aspects,	aspects
	technology,	technology

Si se observan los registros en los que se hacen presentes estas variaciones, la gran mayoría corresponden a un mismo registro por lo que es posible pensar en que un mismo usuario haya ingresado estas etiquetas de manera incorrecta.

Respecto a la utilización de etiquetas ya existentes en el sistema, se observa la presencia de 34 etiquetas que son utilizadas en más de una oportunidad. Dichas etiquetas son utilizadas un total de 109 veces, mientras que las restantes 194 etiquetas son utilizadas una única vez. Es decir, que el total de etiquetas utilizadas en más de una oportunidad alcanza el 36%, siendo el más alto de los tres gestores estudiados.

Etiqueta	Usos	Etiqueta	Usos	Etiqueta	Usos
graphene	10	academic	2	kde	2
java	10	aerogels	2	lv_crossmedia	2
anthropology	9	citation	2	medical	2
imported	6	community	2	mining	2
media	6	concurrency	2	science	2
biomedicine	5	development	2	sensor	2
biomedical	4	diss	2	social	2
data	4	economics	2	software	2
book	3	evolution	2	thermalconductivity	2
myown	3	history	2	v1205	2
toread	3	information	2		
2012	2	infraredmicroscopy	2		

Tabla 7: Etiquetas más usadas en Bibsonomy. Fuente: elaboración propia

Las etiquetas más utilizadas son “graphene” y “java” con un total de 10 usos cada una, y “anthropology” con un total de 9 usos. Se reproduce la Tabla 7 con todas las etiquetas que han sido utilizadas en más de una oportunidad en la muestra.

Métodos de generación y corrección de etiquetas

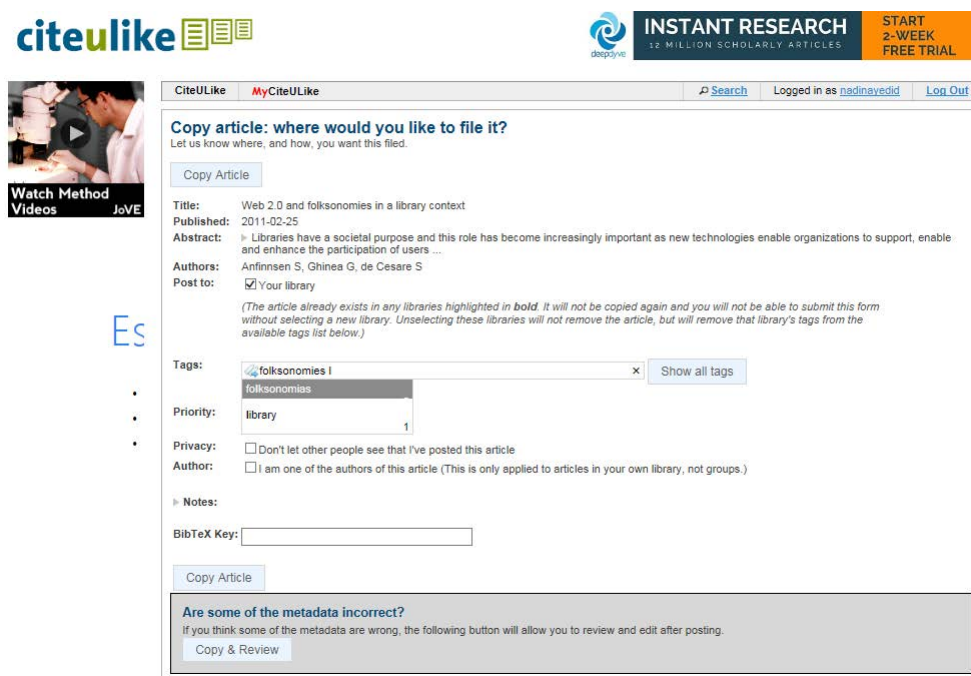
Tal como se ha podido apreciar en los apartados anteriores, existen muchos factores que pueden influir en el tipo, estructura, forma y escritura de las etiquetas en los gestores bibliográficos sociales. Los mecanismos de generación de etiquetas utilizados por cada uno de ellos, comporta un aspecto decisivo en la construcción de la folksonomía del gestor.

A partir de la aplicación de la última sección de la matriz de análisis se pretende conocer cuáles son los métodos de creación y corrección de etiquetas utilizados por cada uno de los tres gestores estudiados.

Métodos de generación y corrección de etiquetas en CiteULike

CiteULike ofrece dos formas de incorporar documentos a la colección personal: copiarlos de otros documentos subidos por otros usuarios del sitio, o bien importar documentos de forma manual desde el propio ordenador. En ambos casos, al momento de guardar dichos documentos en la colección personal, el sistema solicita al usuario que genere las etiquetas que se quiere asociar al artículo para facilitar su identificación y recuperación posterior.

En el primer caso, en el que se copian los documentos de la base que contiene el sistema, se habilita un único cuadro de diálogo, en el que deben ingresarse todas las etiquetas de forma consecutiva. El cuadro de diálogo muestra una (o en ocasiones más) de las etiquetas con las que ya fue etiquetado ese recurso por el usuario del cual se está “copiando” el documento. El nuevo usuario puede elegir entre conservar dichas etiquetas, o bien borrarlas. A medida que se escriben letras en el cuadro de diálogo, el sistema sugiere otras etiquetas que comienzan o contienen las letras



CiteULike **MyCiteULike** [Search](#) [Logged in as nadinevedid](#) [Log Out](#)

Copy article: where would you like to file it?
Let us know where, and how, you want this filed.

[Copy Article](#)

Title: Web 2.0 and folksonomies in a library context
Published: 2011-02-25
Abstract: Libraries have a societal purpose and this role has become increasingly important as new technologies enable organizations to support, enable and enhance the participation of users ...
Authors: Anfinnsen S, Ghinea G, de Cesare S
Post to: ☒ Your library
 (The article already exists in any libraries highlighted in bold. It will not be copied again and you will not be able to submit this form without selecting a new library. Unselecting these libraries will not remove the article, but will remove that library's tags from the available tags list below.)

Tags: [Show all tags](#)
 folksonomias
Priority:
Privacy: ☐ Don't let other people see that I've posted this article
Author: ☐ I am one of the authors of this article (This is only applied to articles in your own library, not groups.)

Notes:
BibTeX Key:

[Copy Article](#)

Are some of the metadata incorrect?
 If you think some of the metadata are wrong, the following button will allow you to review and edit after posting.
[Copy & Review](#)

Figura 1: Ingreso de etiquetas en CiteULike

What tag?

If you've specified keywords fields in your BibTeX file, or KW in your RIS file, then we'll use those as tags. You can also specify a tag name which will be applied to all the articles you import.

Tags:

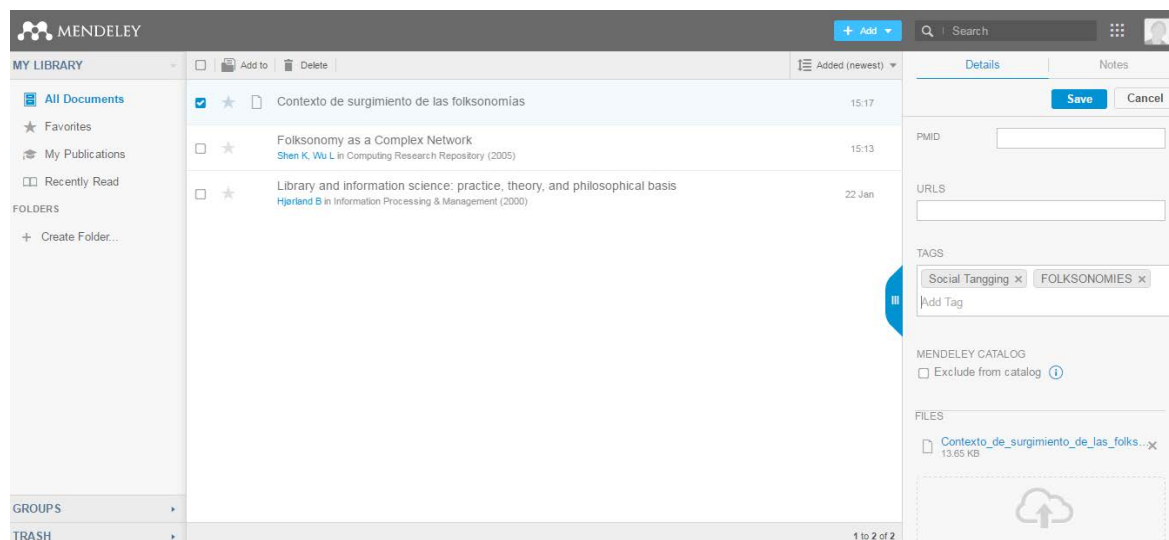
Specify how to parse the keywords or KW field:

- ☒ Make separate tags from all the words
- ☐ Tags are delimited by comma or semicolon; replace whitespace in tags with dashes
- ☐ Don't import any tags

Public or Private?

- ☐ Don't let other people see that I've posted these articles.

Figura 2: Ingreso de etiquetas en CiteULike



MENDELEY [Add](#) [Search](#)

MY LIBRARY [Add to](#) [Delete](#) [Added \(newest\)](#)

Contexto de surgimiento de las folksonomías 15:17

Folksonomy as a Complex Network 15:13
 Shen K, Wu L in Computing Research Repository (2005)

Library and information science: practice, theory, and philosophical basis 22 Jan
 Hjerland B in Information Processing & Management (2000)

DETAILS [Save](#) [Cancel](#)

PMID:

URLS:

TAGS:
[Add Tag](#)

MENDELEY CATALOG: ☐ Exclude from catalog

FILES: 13.65 KB

[Upload](#)

Figura 3: Creación de etiquetas en Mendeley

escritas, mediante un sistema de auto completado. Todas las sugerencias son tomadas de la base de etiquetas propias del nuevo usuario (es decir, otras etiquetas con las que etiquetó otros recursos en el pasado). Además, el usuario tiene también la posibilidad de presionar el botón “Show all tags” para ver todas las etiquetas que él mismo utilizó con anterioridad para describir otros documentos de su colección personal.

En ningún lugar se ofrecen instrucciones respecto de cómo conviene crear las etiquetas, ni se indica que el espacio es usado como delimitador para indicar el fin de una etiqueta y el inicio de otra. Quienes no lo sepan, pueden muy posiblemente ingresar términos compuestos separados por espacios, y encontrarse luego con la novedad de que se han generado dos etiquetas en lugar de una (Figura 1).

Desde esta interface no hay modo de poder visualizar las demás etiquetas que fueron utilizadas para describir este recurso por otros usuarios. Ello solamente puede hacerse en un paso anterior, al visualizar el registro del documento, antes de decidir incorporarlo a la colección personal.

La situación cambia considerablemente en el segundo caso, es decir, cuando se desea guardar documentos importados de forma manual desde el propio ordenador. Tal como puede observarse en la imagen siguiente, en el cuadro de diálogo en el que se deben ingresar las etiquetas se muestra una única “sugerencia”, que consiste en una etiqueta meramente organizativa, que indica la procedencia del archivo (“file-import”) y la fecha en la que fue incorporado al sistema. La posibilidad de ver todas las etiquetas propias, así como la sugerencia de etiquetas a medida que se escribe en el cuadro de diálogo, queda deshabilitada en esta interface. Como contrapartida, se ofrecen mayores indicaciones respecto de cómo ingresar las etiquetas, posibilitando al usuario seleccionar si desea hacer una etiqueta por cada palabra, o si desea reemplazar los espacios por guiones (Figura 2).

En cualquiera de los dos casos, el sistema no detecta mayúsculas de minúsculas, por lo que no importa si el usuario utiliza unas u otras, o una combinación de ambas. El sistema siempre genera las etiquetas en letras minúsculas.

No existe ninguna página de ayuda, o en la cual se den consejos sobre cuáles son los mejores tipos de etiquetas para ingresar.

Dentro de las opciones del perfil privado es posible encontrar una pestaña que permite visualizar todas las etiquetas que uno ha generado, para editarlas o eliminarlas.

Métodos de generación y corrección de etiquetas en Mendeley

El guardado de documentos en la biblioteca personal de Mendeley también puede hacerse o bien copiando documentos que ya forman parte del sitio, o bien importándolos desde el propio ordenador. En ambos casos el procedimiento es el mismo: se guarda el documento sin ingresar ninguna etiqueta para su posterior recuperación. Si el usuario lo desea puede seleccionar el documento que acaba de guardar en su biblioteca personal (sin importar el origen del mismo) y recién allí se habilita una pestaña que permite editar los “detalles” del documento. Al hacer clic sobre esta pestaña se accede a una solapa en la cual es posible ingresar metadata sobre el documento y, como último campo de la descripción, es posible ingresar etiquetas para facilitar su recuperación.

Tal como se observa en la siguiente imagen, las etiquetas deben ingresarse de forma consecutiva en un cuadro de diálogo. El sistema admite mayúsculas y espacios. Para

tags - describe your post

tags

space separated

Please enter valid tags.

recommendation **folksonomies** **folksonomy** **use**

post visibility

viewable for ☒ public ☐ private ☐ other

private documents

upload document Ningún archivo seleccionado

The uploaded documents are stored temporarily. After clicking save, they are added to the post. The documents will be available only to you and members of your groups.

added files

detailed information [\(show all fields\)](#)

Figura 4: Creación de etiquetas en Bibsonomy

dar por terminada una etiqueta e ingresar la siguiente es preciso presionar la tecla “enter” (Figura 3).

En el caso de documentos copiados de la plataforma, el sistema no muestra cuáles son las otras etiquetas con las que el documento fue descrito por los demás usuarios. Esto solo puede verse en el paso anterior, al recuperar el documento.

Sin importar el origen del documento (si fue tomado de la plataforma o si fue importado) el sistema no permite visualizar la nube de etiquetas del sitio al momento de asignar una nueva etiqueta, no se sugieren etiquetas, ni se ofrecen métodos de auto corrección. Tampoco es posible encontrar en el sitio web una página con recomendaciones respecto de cómo generar las etiquetas para maximizar su funcionalidad. No existe la opción de editar una etiqueta que ya ha sido utilizada. La única opción posible es eliminarla del registro y, si se quiere, reemplazarla por otra.

Métodos de generación y corrección de etiquetas en Bibsonomy

Al igual que en los dos casos anteriores, en Bibsonomy también es posible copiar a la biblioteca personal documentos ya existentes en la plataforma o bien importarlos desde el ordenador. Cualquiera sea la opción que se utilice para incorporar un nuevo documento a la biblioteca, al hacerlo el sistema sugiere completar una serie de metadatos sobre el documento, en el cual se incluye un campo para la generación de etiquetas.

Las etiquetas deben ingresarse de manera consecutiva en un cuadro de diálogo. Debajo del mismo puede leerse una frase que indica “space separated” (Figura 4), que da cuenta de que el espacio demarca la finalización de una etiqueta y el comienzo de una nueva. El sistema no distingue entre mayúsculas y minúsculas, por lo que el usuario puede utilizarlas de forma indistinta, pero siempre se generarán etiquetas en letras minúsculas. El sistema no ofrece métodos de auto corrección, y/o auto completado de etiquetas, pero sí muestra recomendaciones de etiquetas para adicionar. Todas las recomendaciones se corresponden con otras etiquetas ya utilizadas por el propio usuario para etiquetar otros documentos (Figura 4).

The screenshot displays the BibSonomy website interface. On the left is a vertical navigation menu with categories: registration/sign-in, basic functions, advanced features, social features, and developers' garden. The main content area is titled 'searching and comparing items.' and includes a 'Table of contents' with links to various tag system topics. Below the table of contents, the section 'How does the tag system work?' explains the difference between folders and tags, providing a practical example of searching for 'politics in France' using both tags. A 'TIP' section at the bottom offers guidelines for creating effective tags, such as using singular words and avoiding unnecessary conjunctions.

registration/sign-in

- Sign in with OpenID account
- Sign in with user account
- Create a new user account
- How to recover your password

basic functions

- User Interface - Overview
- Manage literature and bookmarks
- Use the tag system
- Import Bookmarks
- Bibliography import
- Bibliography export
- Install Browser-Plugin
- Discovery Service

advanced features

- How to change my CV
- Use RSS-Feeds
- Browse publications
- Use the clipboard
- Attach a private document
- URL-Syntax
- (Beta) Addon for Google Docs

social features

- Add friends/contacts
- Group functions
- Discussion: Reviews and Comments
- Use the URL-Resolver

developers' garden

searching and comparing items.

Table of contents

- Use the tag system
 - How does the tag system work?
 - Add tags to bookmarks/publications
 - Edit tags on bookmarks/publications
 - Edit tags via "edit post"
 - Edit tags via "fast edit"
 - Edit tags via "edit tags"
 - Navigation via tags
 - Related tags
 - Similar tags
- System tags
 - Executable system tags
 - Meta system tags
 - Search system tags

How does the tag system work?

From your computer or in real life, you probably know the opportunity to store files/documents in folders and subfolders. A file is placed in a specific folder.

The keyword system works in a completely different way. You can add as many tags as you want to a file/publication. That way you can find all files about a specific topic very fast and easily, for you only have to search for the corresponding tag. The advantage of the tag system is the fact that you can combine different tags when searching to achieve precise and specific results. For example, if you're looking for literature about the topic "politics in France", you can search with both tags "politics" AND "France". As a result, you get all articles that deal with both topics.

With this function, BibSonomy makes it easy for you to store, manage and share your literature with others.

TIP: Good tags are short and concise. In addition, we recommend to always use singular words (e.g. use *tree* instead of *trees* - or even better: use two keywords: *tree* AND *forest*), thus avoiding unnecessary word forms. If you want to use a tag that consists of two words (e.g. *Computer Science*) you should use *PascalCase* (e.g. *ComputerScience*).

Figura 5: Consejos para la creación de etiquetas en BibSonomy

Además, en el caso de documentos copiados desde la plataforma, el sistema permite visualizar las etiquetas con las que fue etiquetado el documento. Sin embargo, no se muestran todas las etiquetas que han sido asignadas a ese documento, sino solo las que ha asignado el usuario del cual se está "copiando" el documento. La cantidad total de etiquetas asignadas al documento solo puede verse desde un paso anterior, cuando se visualiza el registro del documento. En caso de que el usuario lo desee, puede ver la nube completa de etiquetas del sitio web, desde la página de inicio.

Desde la cuenta personal se puede acceder a la pestaña "edit tags", en la cual es posible editar las etiquetas utilizadas para describir los documentos de la biblioteca personal. Las opciones de edición que se presentan son: modificar una etiqueta, añadir una sub-etiqueta, o eliminar una sub-etiqueta.

Finalmente, el sitio web cuenta con una sección de ayuda llamada "getting started", en la cual puede encontrarse un enlace que ofrece información sobre el sistema de etiquetado. Al hacer clic allí se accede a una página en donde se explica cómo se debe hacer para agregar etiquetas a los registros, y también se ofrecen pequeños consejos para generar etiquetas eficientes (Figura 5).

Conclusiones

Se ha observado en los resultados obtenidos que muchas de las etiquetas asignadas por los usuarios responden a los contenidos de los que tratan los documentos, aumentando la usabilidad de las etiquetas por parte de otros usuarios del sistema. En los tres gestores estudiados, las etiquetas de contenido corresponden al mayor porcentaje de la muestra total de etiquetas, sin embargo se observa también un porcentaje considerable de etiquetas refinadoras. Esto resulta altamente llamativo, ya que se trata de etiquetas que por sí solas no tienen poder de descripción para indicar los contenidos de los documentos, sino que deben ser utilizadas en conjunto con otras etiquetas. La presencia de este tipo de etiquetas en los tres gestores da cuenta de "inexperticia" o falta de conocimiento por parte de los usuarios respecto de cómo se deben crear las etiquetas. En este sentido, es posible pensar que la mayoría de las etiquetas

refinadoras no fueron simplemente ingresadas de manera independiente, sino que fueron ingresadas para pertenecer a un conjunto de etiquetas, o a una etiqueta multi-término mal formada. Es por lo tanto altamente probable que esta problemática observada pudiera solucionarse si los usuarios contaran con instrucciones claras y sencillas de cómo debe hacerse el ingreso de las etiquetas, y cómo delimitar el final de una etiqueta y el inicio de otra. Las interfaces de ingreso de etiquetas podrían contar con una simple frase al costado del cuadro de diálogo que informe sobre la forma en que deben ingresarse las etiquetas, tal como “Use the space to separate the tags. Use the hyphen to join two words on the same tag” (“Utiliza el espacio para separar las etiquetas. Utiliza un guión medio para unir dos palabras en una misma etiqueta”).

En casi todos los registros de los tres gestores fue posible encontrar al menos una etiqueta que haga referencia a los contenidos tratados en el documento. Aun así, se puede apreciar que muchas veces estas etiquetas no son ingresadas con ánimos de facilitar el descubrimiento o la recuperación de documentos por parte de otros usuarios en el sitio web. Ello dado que en muchos casos se trata de etiquetas escritas en formas diferentes a las habituales, ya sea mediante el uso de abreviaciones propias o deformaciones profesionales. El uso de estas deformaciones perjudica ampliamente la eficacia de la folksonomía en general, y sería fácilmente desalentado si los usuarios contaran con pequeños consejos relativos a la elección de etiquetas en la pantalla de ingreso de las mismas. Una simple frase en la interface podría recordar a los usuarios del uso colectivo del recurso, como por ejemplo: “Help other users to find this document! Describe it with tags representatives of the content” (“¡Ayuda a otros usuarios a encontrar este documento! Descríbelo con etiquetas que representen su contenido”).

Diferente es el caso de las formas o construcciones lingüísticas utilizadas al ingresar una etiqueta. El uso mayoritario de sustantivos y de frases sustantivadas da cuenta de que existe entre los usuarios por lo menos una “intuición”, cuando no un conocimiento, respecto de cuáles son las mejores etiquetas para describir un recurso en lo que refiere a las formas lingüísticas. Una explicación más detallada respecto del uso de frases sustantivadas para desambiguar un sustantivo simple, o de la poca utilidad de utilizar adjetivos por sí solos, podría tomar más que unas simples líneas en la interface de creación de etiquetas. Por lo que es posible pensar que esta información podría colocarse en una página del sitio web, representada a través de un enlace en la interface de creación de etiquetas. Para que sea útil, el enlace debería estar claramente visible cerca del cuadro de diálogo, y ser explícito respecto de su contenido, incluyendo una frase tal como “Learn how to create good tags” (“Aprende cómo crear buenas etiquetas”). La página enlazada debe contener información clara, sencilla y acotada. Explicaciones largas y complejas podrían atentar contra algunas de las ventajas intrínsecas a las folksonomías (como el bajo nivel de experticia requerido para la creación de etiquetas) y desalentar la lectura de la información por parte de los usuarios.

Es posible pensar que el uso de estas técnicas de “capacitación” de usuarios en la generación de etiquetas podrían complicar un proceso que debería ser, por su propia naturaleza, muy simple. Sin embargo, existen también otras técnicas que podrían redundar en beneficios para el desarrollo de la folksonomía sin generar ningún tipo de *stress* para los usuarios. Particularmente, todas aquellas herramientas que le permitan a un usuario elegir entre otras etiquetas ya creadas en el sistema (en lugar de generar una nueva) o bien contar con funciones de auto-completado, aliviarán la actividad del usuario al mismo tiempo que proporcionarán un medio para eliminar las variaciones de etiquetas analizadas en apartados anteriores de este trabajo. Por ejemplo, es posible que un usuario no ingrese nuevamente en plural un concepto que ya se encuentra ingresado en singular, si el sistema le da la opción de no tener que volver a teclearlo. Así mismo, la herramienta empleada en CiteULike y Bibsonomy

que no reconoce mayúsculas de minúsculas ha probado ser útil para eliminar este tipo de variaciones en las etiquetas.

En función de lo expuesto en los apartados anteriores puede establecerse como conclusión general del trabajo, que el mismo confirma en primera instancia la hipótesis planteada respecto a la utilidad de las folksonomías para la recuperación de información en entornos poco controlados. En función del análisis realizado, se ha observado que muchas de las etiquetas estudiadas son representativas del contenido de los documentos y resultan eficientes para la recuperación de la información por parte de la comunidad de usuarios del sitio web. Se establece además que es posible implementar nuevos métodos (además de los ya utilizados) para mejorar la performance de la actividad de etiquetado por parte de los usuarios.

Referencias Bibliograficas

- » Barité, Mario. 2014. El control del vocabulario en la era digital: revisión conceptual. En *Scire*. Vol. 20, no. 1, 99-108. <<http://www.ibersid.eu/ojs/index.php/scire/article/view/4196>> [Consulta: 1 diciembre 2015].
- » Cañada, Javier. 2016. *Tipologías y estilos en el etiquetado social*. <<http://www.terremoto.net/tipologias-y-estilos-en-el-etiquetado-social/>> [Consulta: 4 enero 2016].
- » Font, F., J. Serrà y X. Serra. 2015. Analysis of the impact of a tag recommendation system in a real-world folksonomy. En *ACM Transactions on Intelligent Systems and Technology*. Vol. 7, no. 1. <<http://mtg.upf.edu/node/3206>> [Consulta: 04 enero 2016].
- » Golder, S.A. y B.A. Huberman. 2006. Usage patterns of collaborative tagging systems. En *Journal of Information Science*. Vol. 32, no. 2, 198-208. <<http://jis.sagepub.com/content/32/2/198.abstract>> [Consulta: 4 enero 2016].
- » Guy, M. y E. Tonkin. 2006. Folksonomies: tidying up tags? En *D-Lib Magazine*. Vol. 12, no. 1. <<http://www.dlib.org/dlib/january06/guy/01guy.html>> [Consulta: 4 enero 2016].
- » Hammond, T.; T. Hannay; B. Lund y J. Scott. 2005. Social bookmarking tools. En *D-Lib Magazine*. Vol. 11, no. 4. <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>> [Consulta: 4 enero 2016].
- » Jäschke, R.; L. Marinho; A. Hotho; L. Schmidt-Thieme y G. Stumme. 2007. Tag recommendations in Folksonomies. En *Knowledge Discovery in Databases*. Vol. 4702, 506-514. <http://link.springer.com/chapter/10.1007%2F978-3-540-74976-9_52> [Consulta: 4 enero 2016].
- » Martínez Tamayo, A. M. y J. C. Valdez. 2009. Indización y clasificación en bibliotecas. Buenos Aires: Alfagrama.
- » Mathes, A. 2004. *Folksonomies: Cooperative Classification and Communication Through Shared Metadata*. <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>> [Consulta: 4 enero 2016].
- » Montero, Y. H. 2006. Indización social y recuperación de información. En *No solo usabilidad*. Vol. 5. <http://www.nosolousabilidad.com/articulos/indizacion_social.htm> [Consulta: 4 enero 2016].
- » Spiteri, L. F. 2007. The structure and form of folksonomy tags: the road to the public library catalogue. En *Webology*. Vol. 4, no. 2, 13-25. <<http://www.webology.org/2007/v4n2/a41.html>> [Consulta: 4 enero 2016].
- » Taylor, A.G. y D.N. Joudrey. 2009. *The organization of information*. 3th ed. Englewood, Colorado: Libraries Unlimited.
- » Weinberger, D. 2007. *Everything is miscellaneous: the Power of the New Digital Disorder*. New York: Times Books.

